Multigrid Optimal Mass Transport for Image Registration and Morphing

Tauseef ur Rehman[a], Allen Tannenbaum[a]
[a]Georgia Institute of Technology, Atlanta, GA, USA

ABSTRACT

In this paper we present a computationally efficient Optimal Mass Transport algorithm. This method is based on the Monge-Kantorovich theory and is used for computing elastic registration and warping maps in image registration and morphing applications. This is a parameter free method which utilizes all of the grayscale data in an image pair in a symmetric fashion. No landmarks need to be specified for correspondence. In our work, we demonstrate significant improvement in computation time when our algorithm is applied as compared to the originally proposed method by Haker et al [1]. The original algorithm was based on a gradient descent method for removing the curl from an initial mass preserving map regarded as 2D vector field. This involves inverting the Laplacian in each iteration which is now computed using full multigrid technique resulting in an improvement in computational time by a factor of two. Greater improvement is achieved by decimating the curl in a multi-resolutional framework. The algorithm was applied to 2D short axis cardiac MRI images and brain MRI images for testing and comparison.

Keywords: optimal transport, elastic registration, mass preservation, multigrid, multi-resolution

1. INTRODUCTION

1.1. Image Registration and Warping

Image registration and warping are amongst the most common image processing problems. Registration is the process of establishing a common geometric reference frame between two or more image data sets. Registration is necessary in order to compare or integrate image data obtained from different measurements. Registration algorithms comprise of a number of steps. Typically, a measure of similarity between data sets is established in order to quantify how close one image is to the other after the application of the transformation. This similarity measure can be based on the pixel intensity values or the proximity of predefined image features such as implanted fiducials, anatomical landmarks, surface contours or ridge lines. The next step is to find the transformation that maximizes the similarity between the transformed and target images. This transformation is commonly constrained to be of a predetermined class (rigid, non-rigid, affine etc.) and found using an optimization framework. Registration has an extensive literature devoted to it with numerous approaches ranging from statistical to computational fluid dynamics. See [3] for a review of methods as well as an extensive set of references.

Image Morphing on the other hand, is a class of techniques that deal with the metamorphosis of one image into another also sometimes known as image interpolation. Given two related images, these techniques generate a sequence of intermediate images in which an image gradually changes into the other over time. There have been a number of different algorithms proposed for image morphing. The common goal in all these algorithms is to find a reasonable warping function between the given two images. Morphing is done by shifting all pixels in the image or a relevant area according to the vector field defined by the warping function. Most of these methods require some kind of feature selection mechanisms. For a general review of image morphing methods, interested readers may refer to [21] and the references therein.

The optimal transport method by Haker et al. is in the class of warping strategies based on continuum and fluid mechanics, in which one tries to use properties of elastic materials to determine deformations. A cost functional (typically quadratic) is defined which penalizes the mismatch between the deforming template and target. In fact, the

Computational Imaging V, edited by Charles A. Bouman, Eric L. Miller, Ilya Pollak
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 6498, 649810, © 2007 SPIE-IS&T · 0277-786X/07/$18

optimal warping map of the L^2 Monge-Kantorovich may be regarded as the velocity vector field which minimizes a standard energy integral subject to the Euler continuity (mass preservation) equation.

1.2. Optimal Mass Transport

The optimal mass transport problem was first formulated by a French mathematician Gasper Monge in 1781, and was given a modern formulation in the work of Kantorovich [20], therefore, is now known as the *Monge-Kantorovich problem*. The original algorithm concerned finding the optimal way, in the sense of minimal transportation cost of moving a pile of soil from one site to another, therefore, the Kantorovich-Wasserstein distance is also commonly referred to as the *Earth Mover's Distance (EMD)*. Optimal transport methods have appeared in econometrics, fluid dynamics, automatic control, transportation, statistical physics, shape optimization, expert systems, and meteorology [3]. The optimal mass transport problem has also been studied within context of imaging and computer vision applications such as similarity metric in shape comparison [4][5] and content-based image retrieval [6] etc.

Recently, Haker et al. have applied the optimal mass transport approach to certain medical image registration problems [1][2]. Rigorous mathematical details for their algorithm are given by Angenent et al. [9][10][11]. There have been a number of algorithms in literature for computing an optimal mass transport but the method by Haker et al. computes the optimal warp from a first order partial differential equation which was computationally an improvement as compared to earlier proposed higher order methods [7][8] and computationally complex discrete methods based on linear programming. In this paper we demonstrate even further improvement in computation time when the same algorithm is applied using multigrid and multi-resolution approach. This method has a number of distinguishing characteristics. It is a parameter free method. It utilizes all of the grayscale data in both images, and places the two images on equal footing. It is thus symmetrical, the optimal mapping from image A to image B being the inverse of the optimal mapping from B to A. It does not require that landmarks be specified. The minimizer of the distance functional involved is unique; there are no other local minimizers. Finally, it is specifically designed to take into account changes in density that result from changes in area or volume.

2. OPTIMAL TRANSPORT THEORY

2.1. Formulation of the Problem

We will briefly provide an introduction to modern formulation of the Monge-Kantorovich problem. We assume we are given, *a priori,* two sub-domains Ω_0 and Ω_1 of $\boldsymbol{R}^d$ with smooth boundaries, and a pair of positive density functions, μ_0 and μ_1 defined on Ω_0 and Ω_1 respectively. We assume that,

$$\int_{\Omega_0} \mu_0 = \int_{\Omega_1} \mu_1 \tag{1}$$

This ensures that we have same total mass in both the domains. We now consider *diffeomorphisms* $\tilde{u}$ from Ω_0 to Ω_1 which map one density to other in the sense that,

$$\mu_0 = \left|D\tilde{u}\right| \mu_1 \circ \tilde{u} \tag{2}$$

which we call the *mass preservation* (MP) property, and write $\tilde{u} \in MP$. Equation (1) is called the *Jacobian equation.* Here, $|D\ \tilde{u}|$ denotes the determinant of the Jacobian map $D\ \tilde{u}$, and $\circ$ denotes composition of functions. It basically implies that if a small region in Ω_0 is mapped to a larger region in Ω_1, then there must be a corresponding decrease in density in order for the mass to be preserved. There may be many such mappings, and we want to pick an optimal one in some sense. Accordingly, we define the squared L^2 Monge-Kantorovich distance as following:

$$d_2^2(\mu_0, \mu_1) = \inf_{\tilde{u} \in MP} \int \left\| \tilde{u}(x) - x \right\|^2 \mu_0(x) dx \tag{3}$$

The *optimal MP map* is a map which minimizes this integral while satisfying the constraint (2). The Monge-Kantorovich functional (3) is seen to place a penalty on the distance the map $\tilde{u}$ moves each bit of material, weighted by the material's mass. A fundamental theoretical result [12][13], is that there is a unique optimal $\tilde{u} \in MP$ transporting μ_0 to μ_1, and that $\tilde{u}$ is characterized as the gradient of a convex function ω, i.e., $\tilde{u} = \nabla \omega$. This theory translates into a practical advantage, since it means that there are no non-global minima to stall our solution process.

In some applications of image registration and image morphing, large intensity discrepancy between corresponding pixels is not desirable. Hence, we can further add a comparison term penalizing change in intensity. The extended energy functional will have the following form:

$$M := C(I_0, I_1, u) + \alpha^2 \int \left\| \tilde{u}(x) - x \right\|^2 \mu_0(x) dx \tag{4}$$

Where C stands for any comparison term, for example squared error, normalized correlation or mutual information. Although the optimal mapping $\tilde{u}$ for this functional is no longer the gradient of a convex function, it can still be solved using the iterative methods proposed in this paper and those proposed by Haker et al. [1][2]. One example of such a functional is as given below:

$$M_\alpha[u] := \int [(I_1 \circ u) - I_0]^2 dx + \alpha^2 \int \left\| u(x) - x \right\|^2 \mu_0(x) dx \tag{5}$$

for a fixed $\alpha \in \mathbf{R}$. Here the first term controls the "goodness of fit" between the (intensity) images I_0: $\Omega_0 \rightarrow \mathbf{R}$ and I_1: $\Omega_1 \rightarrow \mathbf{R}$, and the second Monge-Kantorovich term controls the warping of the map. The function μ_0 is the mass density of the source image defined on Ω_0, and could be the same as I_0 or a smoothed version. It could also be any scalar field that is appropriate for the underlying physical model. Similarly, μ_1 is assumed to be the mass density of the target image defined on Ω_1. By adjusting α, a tradeoff between minimal mass transportation and minimal intensity change is achieved.

2.2. Computing the Transport Map

In this section we outline the various steps involved in computing the optimal transport map using the equivalent method of *polar factorization* as used by Haker et al. [1]. Interested readers are advised to see referenced paper for complete details. In this approach we want to minimize the L^2 Kantorovich-Wasserstein distance functional over mass preserving functions from (Ω_0, μ_0) to (Ω_1, μ_1). This is done by finding an initial mass preserving mapping u: $(\Omega_0, \mu_0) \rightarrow (\Omega_1, \mu_1)$ and then minimizing over $\tilde{u} = u \circ s^{-1}$ by varying s over mass preserving mappings from Ω_0 to Ω_0, starting with s equal to the identity map. If we consider $\tilde{u}$ as a vector field, we can always find a function ω and another vector field χ, with $\mathrm{div}(\chi) = 0$, such that,

$$\tilde{u} = \nabla \omega + \chi \tag{6}$$

i.e., we can decompose $\tilde{u}$ into the sum of a curl-free and divergence-free vector field. Thus, we try to find a mapping s which will yield $\tilde{u}$ without any curl i.e. $\tilde{u} = \nabla \omega$. Once such an s is found we will have $u = \tilde{u} \circ s = \nabla \omega \circ s$, which is the polar factorization of u with respect to μ_0, see [12][14]. As we discussed above, the unique optimal solution of the L^2 Monge-Kantorovich problem has the form $\tilde{u} = \nabla \omega$, and so the problem of finding polar factorization of u and finding the optimal Monge-Kantorovich mapping $\tilde{u}$ are equivalent.

2.2.1. Finding an Initial Mapping

The first step in this algorithm is to find an initial mass preserving mapping. This can be done for general domains using the method of Moser [15][16] or the algorithm proposed in [1]. The later method can simply be interpreted as the solution of a one-dimensional Monge-Kantorovich problem in the x-direction followed by the solution of a family of one-dimensional Monge-Kantorovich problems in y-direction.

2.2.2. Finding the Minimizer

The second step is to adjust the initial mapping found above iteratively using gradient descent in order to minimize the functional M defined in Equation (4), while constraining u so that it continues to satisfy Equation (1). This process iteratively removes the curl from the initial mapping u and, thereby, finds the polar factorization of u. For details on this technique, please refer to [1]. Although the theory has been shown to work for arbitrary dimensions we have implemented and verified the results for the two dimensional case. The overall algorithm is summarized graphically in Figure (1) below.

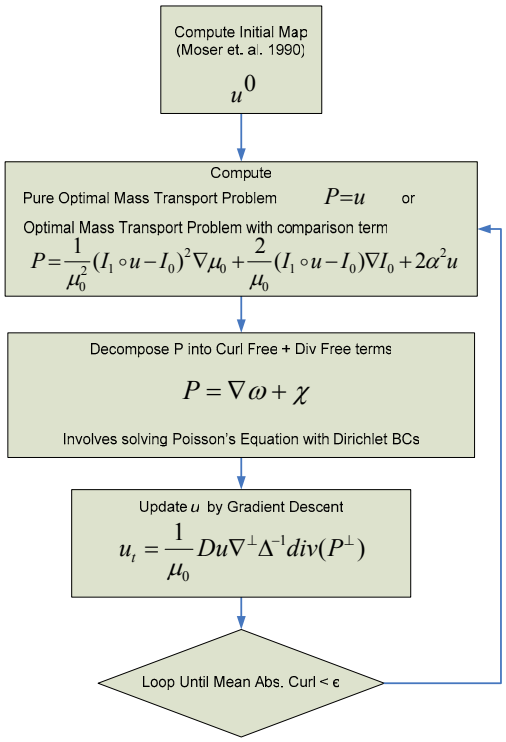


Figure 1. Optimal Mass Transport Algorithm in 2D using Gradient Descent

2.3. Defining the Warping Map

In elastic registration applications, one usually wants to visualize the explicit warping between the two images where one image smoothly deforms to the other. This has been shown to be easily done using the solution from the Monge-Kantorovich problem described above using the following relationship [1]:

$$X(x,t)=x+t(\tilde{u}^*(x)-x) \tag{7}$$

where $X(x, t)$defines our continuous warping map between densities μ_0 and μ_1. Note that when $t = 0$, X is the identity map and when $t = 1$, it is the solution $\tilde{u}^*$ to the Monge-Kantorovich problem.

3. COMPUTATIONAL ENHANCEMENTS

Although the implementation by Haker et al. is computationally an improvement from earlier proposed higher order methods and those based on linear programming it still suffers from the curse of dimensionality and the convergence is very slow. In order to speed up the execution time of the above algorithm shown in figure 1 we analyzed the

computational time of each its constituent elements. This analysis revealed that the solving the Poisson's equation in every iteration was the most expensive operation which was originally done using Matlab's Poisson equation solver which in turn employs sine transforms to invert the Laplacian on a rectangular grid. In order to optimize this calculation we employed the *Full Multigrid* technique which is computationally an optimal method to compute solution of elliptic partial differential equations with *O(N)* complexity, where *N* is the number of mesh points. Multigrid methods overcome the curse of dimensionality by creating a sophisticated fine-to-coarse hierarchy of equation systems with excellent error reduction properties. Low frequencies on the finest grid reappear as higher frequencies on coarser grids, where they can be removed successfully using classical iterative methods. The use of multigrid for this inner loop resulted in an improvement in per iteration (optimal mass transport) computation time by a factor of two. Substantial further improvement in speed of convergence is achieved by decimating the curl in a multi-resolutional framework. These enhancements are described in detail in the following sections.

3.1. Multigrid

The multigrid idea is very fundamental and is the convergence of two separate attempts at finding optimal solutions to iteratively solve linear systems. In one aspect it takes advantage of the smoothing properties of the classical iteration methods at high frequencies (Jacobi, Gauss Siedel, SOR etc) and the error smoothing at low frequencies by restriction to coarse grids. The second predominant view is to regard Multigrid as a preconditioner for Krylov subspace methods such as the Conjugate Gradient method. In this the multigrid's coarse grid restriction procedure implicitly acts as a preconditioning and thereby, reduces the condition number and the associated increase in computational complexity. For a comprehensive overview of multigrid methods, we refer to [17][18]. The basic convergence property of classical iterative methods by which the low frequency Fourier modes of the error decay slowly but the high frequency modes are reduced rapidly is fundamental to the multigrid approach. The essential multigrid principle is to approximate the smooth (low frequency) part of the error on coarser grids. The non-smooth or rough part is reduced with a small number of iterations with a basic iterative method on the fine grid.

The basic components of multigrid methods are discretization, smoothing procedures, coarsening strategy, coarse grid operators, transfer operators from fine grids to coarse and from coarse to fine and finally the cycle type. The particular settings that we used for our multigrid implementation are given below:

3.1.1. Discretization

A discrete Poisson equation with Dirichlet boundary conditions is given as shown below:

$$\Delta_h u_h(x,y) = b_h^{\Omega}(x,y) \qquad ((x,y) \in \Omega_h \tag{8}$$

$$u_h(x,y) = b_h^{\Gamma}(x,y) \quad ((x,y) \in \Gamma_h = \partial\Omega_h \tag{9}$$

where *h* is the coarse grid mesh size. The discretization used for Laplacian (Δ) was standard five-point $O(h^2)$ approximation given by,

$$L_h u_h(x,y) = -\Delta_h u_h(x,y) = \frac{1}{h^2}[4u_h(x,y) - u_h(x-h,y) - u_h(x,y-h) - u_h(x+h,y) - u_h(x,y+h)] \tag{10}$$

Equation (10) above can be formulated as a linear system using lexicographic ordering as $A^h u^h = f^h$, we then apply the multigrid technique to the residual equation, $r^h = f^h - A^h u^h$.

3.1.2. Coarse Grids

The second component of multigrid is the choice of a grid coarsening technique. We chose the most frequently used *Standard Coarsening,* i.e. doubling the mesh size *h* in every direction. The operator *L* also has to be defined on coarser grids. A natural choice for the coarse grid operator L_H is to use the direct analog of L_h on the coarse grid too. This is also known as *Discrete Coarse Grid Approximation* (DCA).

3.1.3. Inter-grid Transfer Operators

The choice of restriction and interpolation operators for the inter-grid transfer of grid functions is closely related to the choice of the coarse grid. The standard transfer operator for *restriction* in the case of standard coarsening is the *Full Weighting* operator which reads as following in stencil notation,

$$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}_{h}^{2h}$$

For the interpolation operator we chose again chose the standard *Bilinear Interpolation* operator which is as given below. Both the operators are applied component wise to each element.

$$\frac{1}{4}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}_{2h}^{h}$$

3.1.4. Smoother

The classical iteration method that we chose for our smoothing operation in the multigrid framework is the *Weighted-Jacobi* scheme. Only one iteration of the smoother was done at each level of the *Full Multigrid* cycle.

3.1.5. Full Multigrid

All iterative methods start with an initial guess at the solution, in our case u^0, and the current error is computed i.e. $e^h = u^h - u^0$ where u^h is the true solution. In case of a linear operator A^h we can solve an equivalent system of equation based on the error and residual given by $A^h e^h = r^h$. Solving this system of equations would give us the desired correction e^h. In the multigrid framework, after the high frequencies of error have been *smoothed* out at one level, the solution is *restricted* to a coarser level and smoothed again. The result is then *interpolated* back to the fine level. This process is recursive in nature and arbitrary number of coarse levels may be used for a cycle in multigrid. *Full multigrid* algorithm is different in the sense that it initializes from the coarsest level. One cycle of this algorithm comprise of the following steps:

a. Compute residual at finest grid.
b. Restrict residual to the coarsest grid
c. Solve by direct method at the coarsest grid
d. Loop from the next-to-coarsest level to the finest level
    i. Interpolate the coarse solution to this level
    ii. Get the r.h.s. for current grid by repeated restrictions
    iii. Do a ***multigrid v-cycle***
    iv. Increment next-to-coarsest level by 1
e. Apply correction to the initial guess

This process is graphically illustrated in the following figure. The thick arrows indicate use of a higher order interpolation scheme instead of the Bilinear Interpolation. This is because when we increment the next to coarsest level by 1 we do not have to use an interpolation complementing the Full weighting scheme to maintain symmetry in a cycle. We used Bi-cubic interpolation for this purpose in order to get greater accuracy.

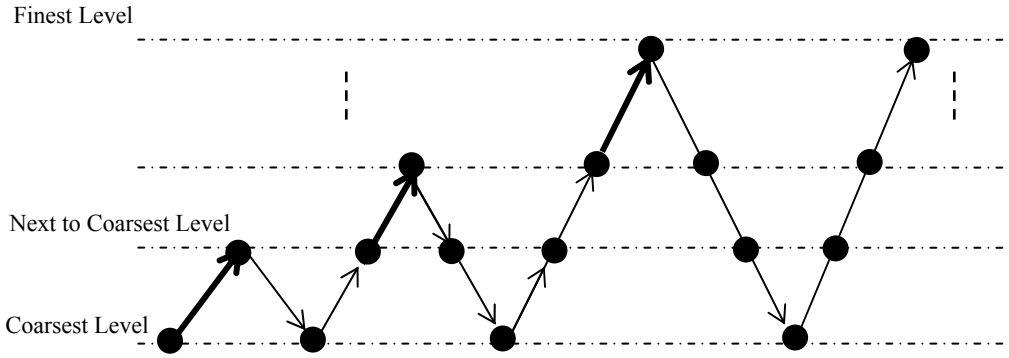


Figure 2. Structure of one Full Multigrid Cycle. The thick arrows indicate use high order interpolation. The upward arrows are for interpolation while the downward arrows are for restriction. The black circles are for an iteration of smoothing operation.

3.2. Multi-resolution Registration

Performing image registration using a multi-resolution approach is widely used to improve speed, accuracy and robustness. The basic idea is that registration is first performed at a coarse scale. The spatial mapping determined at the coarse level is then used to initialize registration at the next finer scale. This process is repeated until it reaches the finest scale. This *coarse-to-fine strategy* greatly improves the registration success rate and also increases robustness by eliminating local optima at coarser scales [19]. Our coarse-to-fine hierarchy comprised of three levels. We used Bi-cubic interpolation to interpolate the solution from the coarse to fine grid. The mean curl of u diminished remarkably fast as compared to the normal case as is shown in figure 3 below. These results are for the brain registration example described in the next section. The figure shows the mean curl of u achieved in the same number of iterations with and without multi-resolution. It should be noted that same number of iteration of multi-resolution complete in almost half the time.

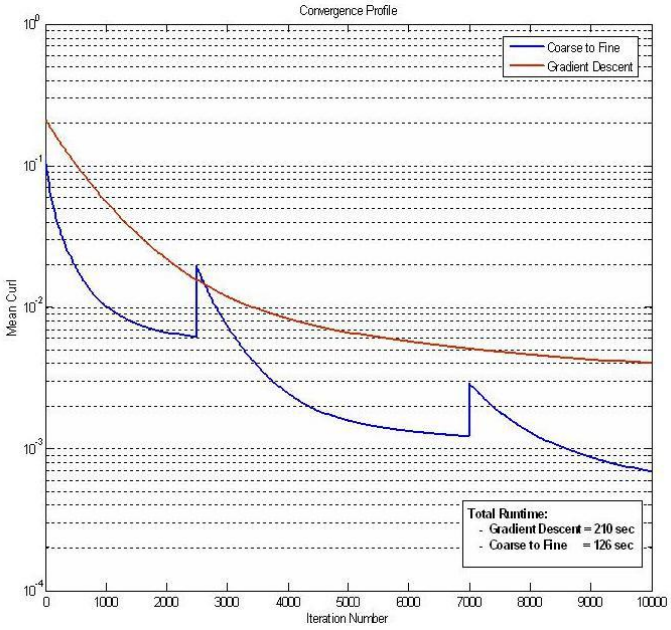


Figure 3. Convergence of the Optimal Mass Transport Registration with and without multi-resolution

4. EXAMPLES

We illustrate our method with two examples using medical images. In the first case we applied our method to two different 2D short axis cardiac MRI images (128×128 pixels) taken from a sequence of images (figure 4). The images were inverted for improved performance. The corresponding deformation grid is shown in figure 5. It can be noted that the deformed grid shows areas of compression of the ventricle. The second example is of brain deformation during surgery, after craniotomy and opening of the dura. The first image is pre-operative and the second is during surgery (figure 6a & 6b). Both images (256×256) were pre-processed to remove the skull. The deformed grid achieved after registration is shown in figure 7. Here also the brain sag can easily be identified from the deformed grid. In both the cases intermediate frames were also found using the warping map described in section 2.3 above. The registration process took approximately 2 minutes for the first example and 5 minutes for the second example on standard 3.2GHz Pentium 4 CPU. The multigrid routines were implemented in C while the remaining algorithm was computed in Matlab.

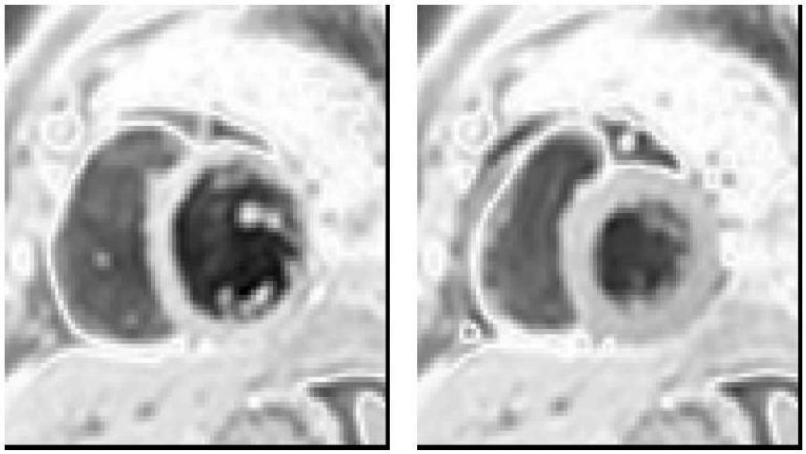


Figure 4. Two frames from a sequence of 2D short axis cardiac images

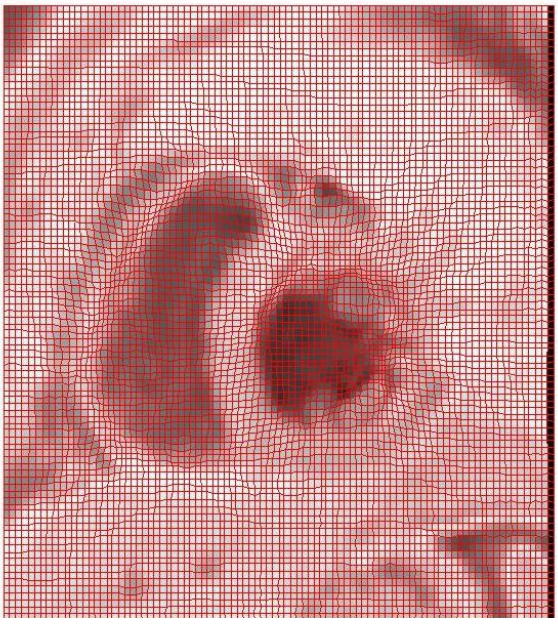
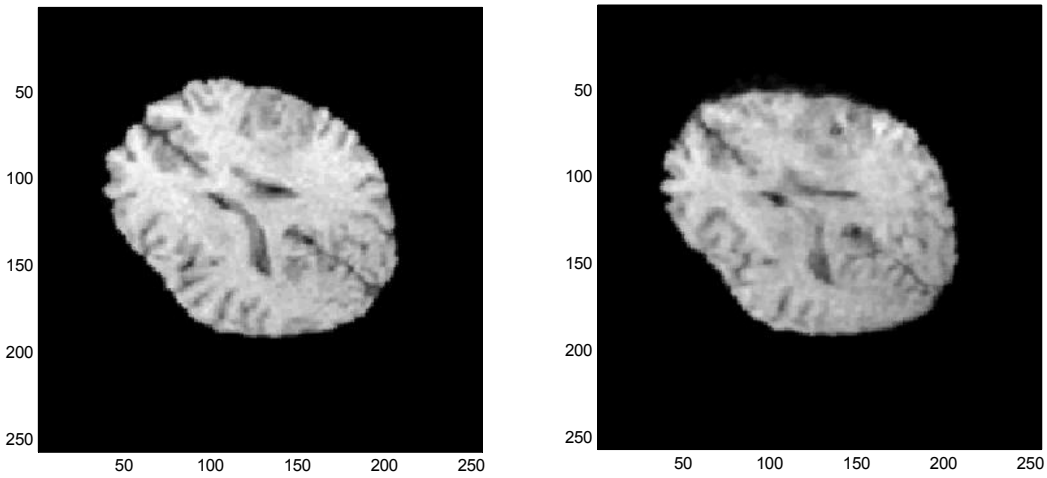


Figure 5. Deformed grid for the cardiac MRI example



(a) (b)

Figure 6. 2D Brain MRI Images. (a) Pre-operative (b) During surgery

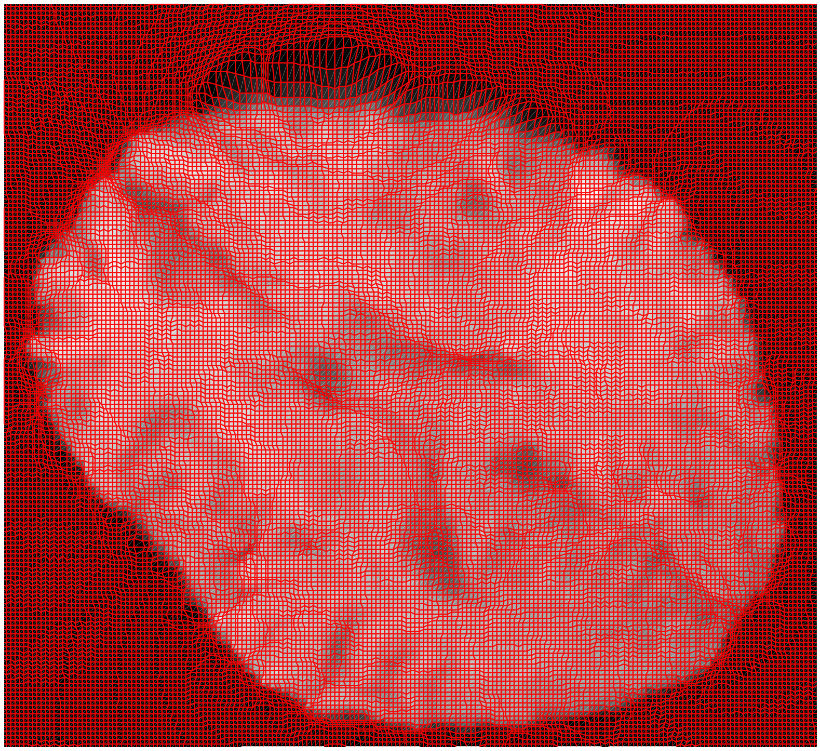


Figure 7. Deformed grid for the brain MRI example

5. CONCLUSION

In this paper, we presented a computationally efficient method for image registration based on the classical problem of optimal mass transportation. We showed that significant improvement in computation time can be achieved by employing multigrid and multi-resolution techniques. The algorithm was applied to some MRI images and reasonable registration results were achieved.

6. REFERENCES

1. S. Haker, L. Zhu, A. Tannenbaum, S. Angenent, *Optimal Mass Transport for Registration and Warping,* International Journal of Computer Vision 60(3), 225-240, 2004
2. S. Haker, A. Tannenbaum, *Optimal Mass Transport and Image Registration,* IEEE Conf. on Variational and Level Set methods in Computer Vision, 2001
3. S. Rachev and L. Rüschendorf, *Mass Tranportation Problems, Vols I and II,* Springer, New York, 1998
4. R. Haralick and L. Shapiro, *Computer and Robot Vision,* Addison-Wesley, New York, 1992
5. D. Fry, *Shape recognition using metrics on the space of shapes,* Ph.D. Thesis, Harvard University, 1993
6. Y. Rubner, C. Tomasi and J. Guibas, *The earth mover's distance as a metric for image reterival,* Technical Report STAN-CS-TN-98-86, Department of Computer Science, Stanford University, 1998
7. G. E. Christensen, R. D. Rabbit and M. Miller, *Deformable templates using large deformation kinetics,* IEEE Trans. On Image Processing, 5:1435-1447, 1996
8. G. E. Christensen, R. D. Rabbit and M. Miller, *A deformable neuroanatomy handbook based on viscous fluid mechanics,* 27th Ann. Conf. on Inf. Sciences and Systems, pp 211-216, 1993
9. S. Angenent, S. Haker and A. Tannenbaum, *Minimizing flows for the Monge-Kantorovich problem,* SIAM J. Math. Analysis, 35:61-97, 2003
10. S. Angenent, S. Haker, A. Tannenbaum and R. Kikinis, *On area preserving maps of minimal distortion,* System Theory: Modeling, Analysis and Control, T. Djaferis and I. Schick (Eds), Kluwer: Holland pp 275-287, 1999
11. S. Angenent, S. Haker, A. Tannenbaum and R. Kikinis, *Laplace-Beltrami operator and brain surface flattening,* IEEE Trans. On Medical Imaging, 18:700-711, 1999
12. Y. Brenier, *Polar factorization and monotone rearrangement of vector-valued functions,* Com. Pure Appl. Math. 64: 375-417, 1991
13. W. Gangbo and R. McCann, *The geometry of optimal transportation,* Acta Math. 177: 113-161, 1996
14. R. McCann, *Polar factorization of maps on Riemannian manifolds,* Geom. Funct. Anal., 11:589-608, 2001
15. B. Dacorogna and J. Moser, *On a partial differential equation involving the Jacobian determinant,* Ann. Inst. H. Poincaré Anal. Non Linéaire, 7:1-26, 1990
16. J. Moser, *On the volume elements on a manifold,* Trans. Amer. Math. Soc. 120: 286-294, 1965
17. W. L. Briggs, V. E. Hensen, S. F. McCormick, *A Multigrid Tutotrial,* SIAM 2000
18. U. Trottenberg, C. Oosterelee, A. Schüller, *Multigrid,* Academic Press, 2001
19. T. S. Yoo, *Insight into Images, Principles and Practice for Segmentation, Registration and Image Analysis,* A. K. Peters Ltd. 2004.
20. L. V. Kantorovich, *On a problem of Monge,* Uspekhi Mat. Nauk., 3:225-226, 1948
21. G. Wolberg, *Image Morphing: A Survey,* Visual Computer, 14:360-372, 1998